The Perception of Place Cues in a Second Language

Mako Fujino

MSc Developmental Linguistics

The University of Edinburgh

2007

Acknowledgements

I would like to thank a number of people who have given me an invaluable help to complete this MSc dissertation project.

First of all, I would like to thank my supervisor Alice Turk for her support, advice, and especially encouragement. I also thank Mitsuhiko Ota for stimulating discussions on second language speech learning throughout the year.

My gratitude goes to everyone who kindly took part in my experiment. Without their help, this project would not have been possible. Special thanks go to technical staffs of the department for helping me setting up the experiments.

Many thanks to my fellow students who have shared countless hours in the lab with me and have given me advice and support.

Finally, I am most grateful to my family for their continuous support and encouragement.

Declaration

I declare that this thesis is of my own work except where explicitly stated in the text. This work has not been previously submitted for any other degree or professional qualification.

Contents Page

List of Tables and Figures

Tables

Figures

Abstract

Many researchers have studied second language (L2) speech perception in terms of the perceptual assimilation into first language (L1) sound categories. However, this approach does not always show whether L2 learners perceive and weigh individual acoustic cues in the same way as native speakers, when they show native-like identification/discrimination of L2 sound contrasts. This issue raises a question especially in the perception of place contrasts which have 2 important cues; a consonant and the adjacent formant transitions. The current study examines if the superficial native-like perception of L2 place contrasts indicates the native-like perception of the individual cues and their optimal weighting, by focusing on the perception of English voiceless labiodental and dental fricatives (/f/-/θ/) by Japanese native speakers. The experimental results showed that though Japanese native speakers did not perceive a fricative cue as accurately as English native speakers, the two groups were not different in terms of cue weighting. The paper discusses the results in terms of methodological and theoretical issues.

1 Introduction

This chapter introduces the overview of the current work. The chapter starts with a statement of focus and aims of the current study. The research questions are presented with brief background issues and how they could contribute to the investigation of larger issues. The chapter finally illustrates a structure of the dissertation.

1.1 Focus of the Study

A number of studies have shown that L2 learners differ from native speakers in various aspects of speech perception. Especially, a failure to pay attention to appropriate acoustic cues causes perceptual problems in L2 learners. For example, it is known that Japanese native speakers have difficulties in discriminating the English /r/-/l/ contrast (Best & Strange, 1992). Iverson and colleagues (2003) found that in the perception of English /r/-/l/ contrast, Japanese native speakers tend to assign greater perceptual weight to the change in F2 than in F3, while English native speakers categorise /r/ and /l/ primarily based on the F3 dimension. Thus, in order for Japanese native speakers to obtain reliable native-like perception of English /r/-/l/, it is necessary to direct their attention to the change in F3 and reduce the perceptual reliance on F2. Flege's Speech Learning Model (SLM) suggests that the ability to learn speech is intact throughout life and L2 learners can actually learn L2 sound categories by shifting their perceptual attention (Flege, 1995).

However, even if L2 learners show the accurate perception of L2 sound contrasts, it does not always indicate native-like use of acoustic cues which could contribute to the perception of contrastive sound categories. For example, Broersma (2005) examined the perception of English word-final voicing contrast /z/-/s/, /v/-/f/, /b/-/p/, and /d/-/t/ by Dutch native speakers whose native language have voicing contrast in the word-initial and word-medial positions but not in the word-final position. The Dutch participants in this study were advanced learners of English. The result of the discrimination test showed that Dutch native speakers distinguished word-final voicing as accurately as English native speakers. However, it was also shown that while the vocalic duration of the preceding vowel was an important cue for the voicing contrast in English native speakers, Dutch native speakers showed less persistent use of the same cue.

Broersma (2005) concluded that native-like identification and discrimination can be obtained without processing acoustic cues in a native-like way. She also pointed out that L2 learners do not care to achieve the native-like use of acoustic cues, especially when the resulting perceptual accuracy meets the communicative needs of L2 learners. Thus, superficial accurate perception of L2 sounds does not always imply native-like processing of acoustic information.

1.2 Research Questions

The discrepancy between readily observable native-like perception and native-like processing of acoustic information raises the following issues in the perception of L2 consonantal place contrasts. Consonantal place contrasts are usually signalled by the consonant itself and the formant transitions in the adjacent vowels and native speakers assign a certain perceptual weight to each cue to determine the percept of the place (e.g., Fischer-Jørgensen, 1972; Nittrouer & Miller, 1997; Walley & Carrell, 1983). The fact that the two important cues are involved in the perception of place contrasts raises two questions in the context of L2 speech perception.

The first question raised by these issues is whether or not the good perception of L2 consonantal place contrasts in consonant-vowel (CV) syllables indicates that L2 learners show as accurate a perception as native speakers for each cue which could contribute to the perception of place contrast. In other words, it is of interest whether L2 learners show the same degree of perceptual sensitivity as native speakers to the consonant itself and formant transitions in the adjacent vowels. For example, in the perception of place contrasts for English stops, a stop burst of 25 ms contains the important information for the place of articulation (Blumstein & Stevens, 1979). However, the percept of place is largely affected by the formant transitions of the adjacent vowel when listeners hear CV syllables with a stop burst and formant transitions signalling different places of articulation (Walley & Carrell, 1983). The same perceptual weighting was found in the perception of French stops (Bonneau, Djezzar, & Laprie, 1996). These studies show that the formant transitions in the adjacent vowels actually serve as a more important cue than the stop burst itself in CV syllables, though native speakers can identify the place of articulation from the stop burst. For fricatives, the role of formant transitions is more variable than in stops depending on place of articulation and the number of fricatives in a language, but it does exert a certain impact on the perception of place contrasts (e.g., Harris, 1958, Wagner, Emestus, & Cutler, 2006).

In the context of L2 speech perception, these studies suggest that L2 learners might be able to show the accurate perception of L2 consonantal place contrasts in CV syllable conditions without having as accurate a perception as native speakers for one of the cues, if they can profit from the other cue. For example, L2 learners might be able to show the accurate perception of English stops in CV syllables without perceiving the stop burst as accurately as English native speakers, if they can profit from the information signalled by formant transitions in the adjacent vowel. Thus, the current study asks whether or not the native-like perception of L2 consonantal place contrasts in CV syllables results from the native-like perception of the individual cues.

The first question raises another question to be addressed. The second question is whether or not L2 learners' perception of individual cues affects the weight assigned to each cue, if L2 learners show the different perceptual sensitivity to individual cues from native speakers. It seems that native speakers assign weight to individual cues according to their informativeness for sound categorisation so that the reliable category perception can be obtained (e.g., Fischer-Jørgensen, 1972; Nittrouer & Miller, 1997). If L2 learners show the different degree of perceptual sensitivity for individual cues from native speakers, it leads on to the fact that informativeness of individual cues for L2 learners might differ from that for native speakers. Then, it would be possible that L2 learners weigh these cues somewhat differently from native speakers.

The two research questions in the current study can be summarised as follows.

1. Does a good perception of L2 consonantal place contrasts in CV syllables indicate

the native-like perception of individual cues?

2. If L2 learners show the different perceptual sensitivity for individual cues from native speakers, does it affect the weighting of these cues?

The current study attempts to investigate these research questions by focusing on the perception of the English voiceless labiodental and dental fricatives (/f/-/θ/) by Japanese learners of English. This case is suitable in the current investigation in terms of the following reasons. First, Japanese native speakers do not have substantial difficulties in the perception of the English /f/-/θ/ contrast in syllable contexts (Lambacher, Martens, Nelson, & Berman, 2001), even if this contrast does not exist in Japanese. Secondly, in the perception of English /f/-/θ/ contrast, both of the consonant and the adjacent formant transitions have certain perceptual importance in English native speakers (Harris, 1958; Nittrouer, 2002). In other words, for the optimal perception, both cues must be reliably perceived and weighted accordingly. The details of acoustic and perceptual characteristics of the English /f/-/θ/ contrast will be discussed in the later sections.

Answering the current research questions could offer some implications in larger contexts. First, it is possible to provide some suggestions on why advanced or even near-native L2 learners have various persistent problems and how these problems might be solved. For example, in consonantal place contrasts, the relative informativeness of a consonant cue increases in front vowel contexts (Fischer-Jørgensen, 1972; Smits, Ten Bosch, & René, 1996) and listeners rely on a consonant cue more in front vowel context (e.g., Mayo & Turk, 2005). That is, in front a vowel context, listeners have to exploit the information from a consonant more accurately in order to identify the place of articulation. In the context of L2 speech perception, if L2 learners achieve a good perception of the place in CV syllables without having as accurate a perception as native speakers for a consonant cue, it might be the case that their perception is more vulnerable in front vowel contexts than in non-front vowel contexts. Thus, by examining the perception of individual cues, it might be possible to predict specific contexts where L2 learners have persistent specific contexts where L2 learners have persistent problems. If this is the case, the findings could contribute to developing effective trainings for L2 learners by which the perception of a specific acoustic dimension can be improved.

Secondly, the current study could contribute to the investigation of perceptual cue integration mechanisms in general. A number of speech perception studies have shown that context-dependent perceptual informativeness of individual cues does matter in determining the relative weight assigned to each cue (e.g., Fischer-Jørgensen, 1972; Nittrouer & Miller, 1997). However, in the study of non-speech auditory perception (Holt & Lotto, 2006) and cross-modal (visual and auditory) perception (Battaglia, Jacobs, & Aslin, 2003), it was found that cue weighting might also be influenced by global informativeness of individual cues determined by the whole perceptual experience of perceivers (Holt & Lotto, 2006). In other words, if cue A is likely to be more informative than cue B in many of the perceptual situations, cue A might still be weighted even under a specific context where cue B is much more informative than cue A. Holt and Lotto (2006) referred to cue A in this case as *salient*. In the context of the current study, if L2 learners' cue weighting cannot be predicted from the perceptual informativeness of individual cues in a specific context, there is a possibility that the salience of these cues determined by learners overall perceptual experience (L1 and L2 experience) might somehow override the local relative informativeness of these cues. The current study does not attempt to offer direct evidence for this issue, but in a larger context, it would be an issue to be addressed.

Thirdly, the current study could contribute to the discussion on whether or not the perception of individual cues is an important issue in the study of L2 speech perception. Vallabha (2006), for example, argues that the perception of the overall acoustic configuration matters more than the perception and the weighting of individual cues. Although the current study is based on the assumption that the perception and the weighting of individual cues matters, the results should also be discussed in the light of the alternative approach.

Thus, the current study could contribute to the study of L2 speech perception and, possibly, to the study of speech perception in general.

1.3 Structure of the Dissertation

The current work is an experimental study which attempts to find answers to the two research questions discussed in the previous sections. Chapter 2 discusses the previous literature relevant to the current study. First, it is discussed what kind of acoustic cues can signal place contrasts and how native speakers exploit them in the optimal way. Secondly, the influential model of L2 speech perception and its problems are presented and the following section discusses the alternative approach. Thirdly, the details of acoustics and perception of the English /f/-/θ/ contrast is discussed. Chapter 3 presents three experiments followed by statistical analyses and discussions of the results. Chapter 4 examines the experimental results in the light of the initial research questions and discusses their implications. Finally, chapter 5 consists of conclusions and implications for future research.

2 Backgrounds

This chapter summarises the previous literature relevant to the current study. Section 2.1 discusses studies on the perception of place contrasts. The following section 2.2 takes a close look at issues in L2 speech perception. Finally, section 2.3 discusses the acoustics and perception of the English /f/-/θ/ contrast which will be examined in the current study.

2.1 The Perception of Place Contrasts

2.1.1 Acoustic Cues in Speech Sounds

Speech sound categories are acoustically distinguished by multiple cues. In the well studied case, Lisker (1986) found that the English /b/-/p/ contrast can be characterised at least by 16 acoustic dimensions spreading from closure to post-closure of the stop articulation; duration of closure, F0 contour in pre-closure, voice onset time (VOT) in post-closure, among other things.

However, even if acoustic analyses show multiple dimensions for a phonetic contrast, it does not mean that listeners exploit all of these acoustic dimensions equally to perceive the phonetic contrast. Some of the acoustic dimensions are potentially more important than others for phonetic contrasts (Repp, 1982). Listeners weigh acoustic dimensions so that the crucial elements for phonetic contrasts can be perceived. For example, for the English /b/-/p/ contrast, VOT is considered to be a strong cue (Lisker & Abramson, 1970). However, it is also true that other cues do have a certain role in the perception of the English /b/-/p/ contrast (Lisker, 1986). It can be claimed that the reliable speech perception is achieved by listener's ability to utilise multiple acoustic cues each of which could be imperfect of its own but forms a complex speech sound category all together (Diehl, Lotto, & Holt, 2004).

2.1.2 Cues for Place Contrasts

A number of studies have shown that place contrasts are signalled by a consonant itself and formant transitions in the adjacent vowel. This claim is held across different manners of articulation; stops (Walley & Carrell, 1983), fricatives (Harris, 1958; Heinz & Stevens, 1961), and nasals (Kurowski & Blumstein, 1984; Malécot, 1956). The relative weight assigned to each cue differs across and within different manners of articulation.

In principle, acoustic cues with bigger differences are more informative for listeners and are likely to be paid greater weight (Holt & Lotto, 2006). Three important factors which determine the informativeness of each cue are discussed here. First, relative informativeness of acoustic cues is context-dependent. For example, in the perception of stop place contrasts, vowel contexts affect the relative cue weighting for a stop burst and formant transitions. In general, formant transitions following the different places of articulation shows greater differences in non-front vowel contexts than in front-vowel contexts. In other words, formant transitions are more informative in non-front vowel contexts. As a consequence, the relative weight assigned to formant transitions increases in non-front vowel contexts, while it decreases in front vowel contexts (Fischer-Jørgensen, 1972; Smits, Ten Bosch, & René, 1996). Thus, when multiple cues are available for the perception of phonetic contrasts, the perceptual system of native speakers weighs more informative cues depending on vowel contexts.

Two other factors which affect the relative informativeness of acoustic cues are the acoustic properties of a consonant and the number of place contrast that a language has. These effects seem to be particularly important in fricatives (e.g., Wagner et al., 2006) and will be thoroughly discussed in the following section.

2.1.3 Cues for Place Contrasts in Fricatives

For the perception of fricative place contrasts, a fricative itself and formant transitions in the adjacent vowel were found to be the two important cues (e.g., Harris, 1958; Heinz & Stevens, 1961). The fricative cue can be decomposed into smaller cues, such as spectral features (Hughes & Halle, 1956), duration (Jongman, 1989), and amplitude (Behrens & Blumstein, 1988; Heinz & Stevens, 1961) and each of which is considered to have a different degree of perceptual importance. However, in the current study, the details of individual properties contained in the fricative are not thoroughly discussed. Instead, the discussion focuses on the relative role assigned to a fricative and formant transitions in the adjacent vowel.

In fricative place contrasts, besides the context effect discussed in the previous section, there are two other important factors which affect the relative cue weighting of the two cues. The first factor is acoustic differences between sibilants (e.g., alveolar, palato-alveolar) and non-sibilants (e.g., labiodental, dental). Sibilants are characterised by distinct spectral shapes, while non-sibilants have similar spectral properties (Behrens & Blumstein, 1988; Heinz & Steven, 1961; Hughes & Halle, 1956; Stevens, 1960). In other words, the spectral shape of sibilants is highly informative, whereas that of non-sibilants is less informative. The perception studies have revealed that English sibilant /s/ and /ʃ/ are reliably discriminated by a fricative cue alone, but the perception of non-sibilant /f/ and /θ/ are largely based on formant transitions (Harris, 1958). Thus, the relative informativeness of a fricative and formant transitions is partly determined by acoustic characteristics of sibilants and non-sibilants.

However, acoustic properties of sibilants and non-sibilants are not enough to determine the way of acoustic cue weighting, because the optimal cue weighting depends on the sound inventory of each language. When a language has a three-way contrast for sibilants, for example in Polish (Nowak, 2006) and Shona (Bladon, Clark, & Mickey, 1987), the relative importance of formant transitions increases, because three sibilants have a similar spectral shape. In other words, the relative informativeness of fricatives decreases because of the more crowded inventory. On the other hand, Wagner and colleagues (2006) found that the perception of a non-sibilant /f/ by German and Dutch native speakers does not exploit formant transitions as much as English native speakers do because these languages do not have the /f/-/θ/ contrast. In these languages, /f/ does not have to be discriminated from /θ/. As a result, the relative informativeness of a non-sibilant fricative cue increases because it has to be distinguished only from sibilants. Thus, the sound inventory of language is highly important in determining relative cue weighting in fricatives.

2.1.4 Summary

Section 2.1 discussed what kind of acoustic cues signal place contrasts and how native speakers exploit them in the optimal way. Consonantal place contrasts are generally signalled by a consonant and the adjacent formant transitions. The listeners weigh these cues according to how informative each cue is in signalling the contrastive places. In fricatives, the optimal cue weighting mainly depends on 1) vowel contexts, 2) acoustic properties of fricatives, and 3) a fricative inventory.

2.2 Second Language Speech Perception

2.2.1 Speech Perception through First Language Filter

Section 2.1 illustrated how native speakers achieve the optimal perception of contrastive sound categories with the particular focus on fricative place contrasts. When it comes to L2 speech perception, various sources suggest that the optimal perception is not readily available to L2 learners. It is known that speech perception ability starts to be reorganized to fit one's native language even within the first year of life (e.g., Kuhl, 1991; Werker & Tees, 1984) and the optimal speech perception for one's native language develops during the childhood (e.g., Hicks & Ohde, 2005; Hirai, Yasu, Arai, & Iitaka, 2005; Mayo & Turk, 2004, 2005; Nittrouer, 2002, 2004; Nittrouer & Lowenstein, 2007; Nitteour & Miller, 1997; Wardrip-Fruin & Peach, 1984). This development of speech perception allows native speakers to reliably perceive the contrastive sound categories in their native language. However, as a consequence of this perceptual development, non-native/L2 speech sounds come to be perceived through a native-language filter, or even through a native-dialect filter (Dufour, Nguyen, & Frauenfelder, 2007).

The Perceptual Assimilation Model (PAM) has been an influential model of L2 speech perception (Best, 1995). The model suggests that L2 sounds are perceived in reference to phonetic categories of listener's native language. Best (1995) argues that the relationship between L1 and L2 sounds are based on gestural similarities and the degree of similarity determines how L2 sounds are assimilated into L1 sound categories, at least at the beginning of L2 learning.

According to Best (1995), the perception of non-native sounds can be described by three types of perceptual assimilation. First, a non-native sound is assimilated to listener's native sound category. The degree of assimilation differs depending on whether the non-native sound is perceived as a good or poor exemplar of the native category. Secondly, a non-native sound is perceived as a speech sound but cannot be categorised into any of the native categories. Thirdly, a non-native sound is heard as non-speech sounds and therefore cannot be assimilated into any of the native categories.

When two non-native sounds are assimilated into two distinct native sound categories, the discrimination between these non-native sounds is considered to be good. For example, Best (1990) found that English-speaking adults showed a very good discrimination for the Ethiopian ejective contrast /p'/-/t'/, which does not exist in English. The discrimination between these two non-native sounds is good because they are considered to be assimilated into English /p/ and /t/ respectively.

On the other hand, if two non-native sounds are assimilated into a single native sound category, the discrimination is generally poor. For example, it is well known that Japanese-speaking adults have difficulties in discriminating the English /r/-/l/ contrast. In this case, it is considered that these two sounds are assimilated into a single liquid category /r/ in Japanese (Best & Strange, 1992). The discrimination of non-native sounds is problematic also when the assimilation of non-native sounds overlaps. For example, Aoyama (2003) found that the discrimination of the English word-final /n-/ŋ/ was problematic for Japanese native speakers because Japanese do not have /ŋ/ in the word-final position. While English word-final /n/ is mainly assimilated into Japanese /N/, English word-final /ŋ/ is primarily assimilated into /Ngɯ/ but also into /N/. These assimilations cause a problem in the discrimination of the English word-final /n/-/ŋ/ because the assimilation of English /n/ and /ŋ/ is partly overlapped in Japanese /N/. Thus, the lack of equivalent phonological categories between native and non-native languages causes problems in the perception of non-native sounds.

PAM's account is not only limited to a phonological level but also exerts influences on a phonetic level (Best & Tyler, 2007). For example, French has /w j r l/ as same as English, but their phonetic realisation is not necessarily the same in the two languages, especially for /r/ which is realised as a central approximant in English and as a uvular fricative in French. Because of this phonetic difference, French /r/ is not very well assimilated into English /r/ and French native speakers have some difficulties in the perception of English /r/ (Hallé, Best, & Levitt, 1999). Another example is the perception of Thai word-final stops by English native speakers. Although Thai and English have the same set of stops in the word-final position, Thai word-final stops are invariably unreleased (Tingsabadh & Abramson, 1993) and English word-final stops are variably released (Byrd, 1993). Tsukada (2006) examined the perception of Thai word-final stops by English native speakers and found that they showed lower perceptual accuracy for Thai word-final voiceless stops /p t k/. The result suggests that L2 sounds cannot be mapped onto phonologically equivalent L1 sound categories, if their phonetic realisation differs.

Thus, the various phenomena in L2 speech perception have been accounted for in the framework of PAM proposed by Best (1995). However, there are some problems with this model. For example, the criterion for a particular sound to be considered as "categorised" as a specific L1 category is somewhat arbitrary. Harnsberger (2001) suggested that a non-native sound should be considered to be clearly "categorised" as a native category if the listeners identify the non-native sound as a particular native sound more than 90% of the time. Nevertheless, there is no uniform definition on how to decide whether or not a non-native sound is considered to be "categorised" as a native sound. Consequently, it seems that PAM is often used as a post-hoc explanation. That is, perceptual assimilation is rarely tested directly by thoroughly examining the gestural similarities and dissimilarities between L1 and L2 sounds. As a result, the relationship between L1 and L2 sounds are induced from how L2 learners actually perceive L2 sounds. Thus, it is not always clear what is actually happening in L2 speech perception. The next section discusses one of the recent approaches which could add new insights into L2 speech perception study.

2.2.2 Recent Approach to Second Language Speech Perception

One of the recent approaches attempts to explain L2 speech perception by examining whether or not L2 learners use acoustic cues in the same way as native speakers. For example, Japanese speakers' inability to perceive the English /r/-/l/ contrast was attributed to single-category assimilation in PAM's account (Best & Strange, 1992). However, the study by Iverson and colleagues (2003) revealed that the Japanese learners' problem resided in the weighting of F2 and F3 dimensions. They found that Japanese native speakers weighted F2 more than necessary, while they failed to weigh the F3 which is crucial for the perception of the English /r/-/l/ contrast.

Furthermore, this approach has revealed subtler differences between native speakers and L2 learners. For example, Ingram and Park (1997) found that Japanese native speakers were able to identify the English /i/-/I/ contrast very accurately. However, Japanese native speakers primarily rely on the durational cue for the English /i/-/I/ contrast and they cannot use the spectral cue as accurately as English native speakers (Morrison, 2002). Another example comes from the perception of the English word-final voicing by Dutch native speakers. Broersma (2005) found that Dutch native speakers distinguished the word-final voicing as accurately as English native speakers. However, she also found that Dutch native speakers differed from English native speakers in that they did not use the vocalic duration of the preceding vowel as an important cue for the voicing contrast.

Thus, the important message from these studies is that difference between native speakers and L2 learners might become clear only when the use of specific acoustic dimensions is examined. As already discussed in chapter 1, this finding raises issues in the perception of place contrasts, because the perception of place contrasts relies on at least two important cues and these cues must be weighted in the optimal way.

2.2.3 Summary

Section 2.2 discussed how L2 speech perception differs from L1 speech perception. The Perceptual Assimilation Model (Best, 1995) has attributed the reason to how

L2 sounds are assimilated into L1 sound categories. The recent studies advanced the investigation further by examining how L2 learners differ from native speakers in terms of the use of various acoustic cues. One of the important findings is that the native-like identification/discrimination of L2 sounds does not imply the native-like use of acoustic cues.

2.3 English Voiceless Labiodental and Dental Fricatives

2.3.1 The Perception by English Native Speakers

Section 2.3 discusses the English /f/-/θ/ contrast which is used in the current study to examine the research questions. English non-sibilant /f/ and /θ/ are often confusing even for English native speakers (Jongman, Wang, & Sereno, 2000). It is also known that the perception of these sounds shows greater influence from semantic and facial information because of their acoustical similarities (Jongman, Wang, & Kim, 2003). Acoustically, /f/ and /θ/ have similar energy distribution and do not show any substantial spectral differences (e.g., Behrens & Blumstein, 1988; Stevens, 1960). An acoustic study by Jongman and colleagues (Jongman, Wayland, & Wong, 2000) also found that /f/ and /θ/ in American English cannot be reliably distinguished by their spectral peak location though voiced counterparts /v/ and /ð/ show distinct spectral peak locations. It is important to note that the majority of these studies were conducted in General American English. For English /θ/, it is known that some dialectal differences exist in its realisation. While interdental realisation is more common in General American English, dental realisation is predominant in Southern British Standard English (Ladefoged & Maddieson, 1996) which will be used in the current study. In general, /θ/ with dental realisation and /f/ is slightly less confusing than /θ/ with interdental realisation and /f/ (Jones, 2005).

Despite similar acoustic features between /f/ and /θ/, English native speakers seem to be able to identify these fricatives to some extent by isolated fricatives. Jongman (1989) found that /f/ is correctly identified by 50 ms of its noise portion and /θ/ is also correctly identified when the full noise portion is presented, though the identification score did not reach 100%. Although spectral shapes might not reliably signal the difference, the listeners seem to be able to profit from small acoustic differences to identify these fricatives.

However, in reality, the English /f/-/θ/ contrast is likely to be signalled not only by noise frication but also by formant transitions in the adjacent vowel. The formant transitions of /f/ and /θ/ are highly informative in that both F2 and F3 move to different directions; both F2 and F3 show rising transitions for /f/, while both show falling transitions for /θ/. A number of studies have shown that the percept of English /f/ and /θ/ can be obtained without actually presenting /f/ and /θ/ fricatives. When CV syllables consisting of a neutral noise and a vowel with either /f/-transition or /θ/-transition, English native speakers can identify these syllables either as /f-vowel/ or /θ-vowel/ based on the information from the formant transitions (Carden, Levitt, Jusczyk, & Walley, 1981). It was also found that a vowel portion truncated from /f-vowel/ or /θ-vowel/ syllables resulted in significantly better identification of consonants (55.0-87.5%) than that truncated from /s-vowel/ or /ʃ-vowel/ syllables (0.0-22.5%) (LaRiviere, Winitz, & Herriman, 1975).

When both cues are available, it is considered that formant transitions in the adjacent vowel do have an important role in the perception of /f/ and /θ/, at least compared to those in sibilants (Heinz & Stevens, 1961). Importantly, when the places signalled by a noise and formant transitions are conflicting between /f/ and /θ/, the perception is largely affected by the information from formant transitions (Harris, 1958). However, it should also be noted that the perception of the /f/-/θ/ contrast cannot be completely determined by formant transitions. Nittrouer (2002) carried out cue weighting study with stimuli created by combining very similar natural /f/ or /θ/ noise with a synthetic vowel continuum whose two endpoints have appropriate formant transitions either for /f/ or /θ/. Even though natural /f/ and /θ/ used to create the stimuli were acoustically very similar, the identification function for the /f/-continuum and the /θ/-continuum was clearly separated. In other words, the listeners weighted noise frication to some extent. Thus, though formant transitions exert an important influence to the perception, fricatives also convey a certain amount of place information for /f/ and /θ/.

2.3.2 The Perception by Japanese Native Speakers

Japanese has neither /f/ nor /θ/. This sound contrast is, therefore, novel to Japanese learners of English. Some studies have shown how these sounds are perceived by Japanese native speakers when they are adjacent to vowels. Lambacher and colleagues (2001) examined the perception of English voiceless fricatives /f, s, ʃ, θ, h/ by Japanese native speakers and English native speakers by presenting five English fricatives with various vowels /i, ε, a, o. u/ in CV, VCV, VC contexts. The English proficiency of the Japanese participants is considered to be intermediate. The participants performed a five-forced-choice task by choosing the fricative sound they heard from "F, S, SH, TH, H". The result showed that Japanese native speakers identified /f/ and /θ/ less accurately than English native speakers. However, Japanese native speakers (9.92%) perceived /f/ as /θ/ less frequently than, or as frequently as English native speakers (11.11%). On the other hand, the lower identification of /θ/ mainly resulted from the confusion of /θ/ as /s/ (28.23%) and not as /f/ (12.80%). In fact, /θ/ tended to be perceived as /s/ twice as frequently as /f/. A multidimensional scaling analysis for the overall confusion matrix is shown in Figure 1. It seems that Japanese native speakers did not have a substantial difficulty in discriminating the English /f/-/θ/ contrast.

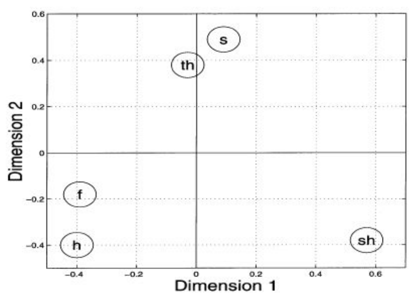

Figure 1. Spatial relationship of English fricatives in Japanese speakers

(Lambacher et al., 2001, p. 340)

Researchers have explained the reason why Japanese native speakers do not have substantial difficulties with the English /f/-/θ/ contrast in the framework of PAM

(Best, 1995). The reason is attributed to the fact that /θ/ is likely to be assimilated to Japanese /s/ and /f/ into /h/ whose allophone is [ɸ] (Lambacher et al., 2001). It was also found that English /θa/ was perceived as a poor member of Japanese /sa/ by Japanese native speakers (Guion, Flege, Akahane-Yamada, & Pruitt, 2000). According to their study, this confusion between English /θa/ and Japanese /sa/ was persistent even in advanced Japanese learners of English who had lived in the United States for 3 years and 1 month on average (Guion et al., 2000). One of the reasons why English /sa/ is persistently perceived as Japanese /sa/ might reside in the phonetic realisation of Japanese /s/. It is said that Japanese /s/ is realised by dental-laminal articulation and is acoustically distinct from English /s/ (Li, Edwards, & Beckman, 2007). In other words, Japanese /s/ is closer to English /θ/ than English /s/ is. Because of this phonetic realisation of Japanese /s/, English /θ/ might be persistently assimilated into Japanese /s/ even in advanced learners.

In summary, English /f/ and /θ/ is likely to be assimilated into two different Japanese categories and, therefore, Japanese native speakers do not have serious problems with discriminating these two non-native sounds. However, as discussed in the previous section, this explanation by PAM is somehow post-hoc and it does not show how exactly Japanese native speakers achieve the perception of the English /f/-/θ/ contrast. Importantly, it is not clear if Japanese native speakers can accurately perceive the individual cues contributing to the perception of the English /f/-/θ/ contrast in the syllable condition. The current study attempts to examine this issue by looking into how L2 learners use each acoustic cue relevant to the perception of the English /f/-/θ/ contrast.

2.3.3 Summary

English /f/ and /θ/ are acoustically similar and perceptually confusing even for English native speakers. Formant transitions, therefore, play an important role in the perception of the /f/-/θ/ contrast in the syllable conditions. Japanese native speakers do not have a substantial problem with this contrast in syllable contexts, though it does not exist in Japanese. The current study attempts to have a closer look at this case by examining the perception of individual cues involved in the perception of the English /f/-/θ/ contrast.

3 Experiments

Chapter 3 presents the experiments carried out for the current study. A design of the experiments is briefly presented first and is followed by the details of each experiment.

3.1 Design

The current study was designed to test the two research questions by examining the perception of the English /f/-/θ/ contrast by Japanese native speakers. The vowel context is limited to /ɑ/ in this study and the vowel-context effect is not considered. The two research questions are repeated here in the context of this specific sound contrast and language groups.

1. Does a good perception of the English /f/-/θ/ contrast by Japanese native speakers in /fɑ/-/θɑ/ syllables indicate their native-like perception of each cue; (a) a fricative, and (b) formant transitions?

2. If Japanese native speakers show the different perceptual sensitivity for a fricative and formant transitions from English native speakers, does it affect the weighting of these cues?

The current study consists of three experiments. Experiments 1 and 2 examine the perception of individual cues (fricatives, formant transitions) by Japanese and English native speakers. The results from these studies show if Japanese native speakers have the native-like perception of individual cues contributing to the perception of the English /fɑ/-/θɑ/ contrast. Experiment 3 examines the perception of the English /fɑ/ and /θɑ/ syllables and looks into how the two cues are weighted by Japanese and English native speakers.

The participants took part in Experiment 1 first and they came back for Experiments 2 and 3 approximately 2 weeks after Experiment 1. Experiments 2 and 3 were conducted in the same experimental session. The order of the two experiments was counter-balanced across the participants.

3.2 Experiment 1: The Perception of Fricatives

The purpose of Experiment 1 was to examine if Japanese native speakers show the native-like perception of /f/ and /θ/ fricatives.

3.2.1 Participants

Eighteen (4 male, 14 female) Japanese native speakers (mean age = 33.3, range = 22-50) and 18 (7 male, 11 female) English native speakers (mean age = 24.3, range = 20-36) living in Edinburgh participated in the experiment[1]. All reported normal speech and hearing. All of the Japanese participants grew up in a Japanese-speaking family, primarily in Japan. Among the Japanese participants, the length of residence in English speaking countries varies from 4 months to 20 years (mean length = 3 years 10 months). All of them have been working or studying using English and are considered to be advanced learners of English. Most of the Japanese participants started studying English at the age of 12 or 13 years old when English teaching starts at school in Japan. There were four Japanese participants who had earlier contact to English. Two of them started to learn English earlier in Japan either from 6 or 10 years old. The other two lived in the United Kingdom for 2 years and 6 months or 3 years and 6 months during their childhood, but they constantly used Japanese in their everyday life. The English participants were speakers of Southern British Standard English. The participants were paid for their participation.

3.2.2 Materials

Three /fɑ/ and three /θɑ/ syllables were recorded by a female speaker of Southern British Standard English with normal speech and hearing. The recording was carried out in the recording studio in the Department of Linguistics and English Language, the University of Edinburgh. The recordings were digitised at 22.05 kHz with

[1] As the mean age of the two groups was largely different, it was ensured that the experimental results were not correlated to participants' age.

16-bit quantisation and were exercised using Adobe Audition and Praat. The six tokens were amplitude-normalised based on the single peak amplitude of a vowel. In order to ensure the quality of the stimuli, five native speakers of Southern British Standard English performed an identification test on the recorded syllables. All of the recorded syllables were correctly identified by the listeners.

Three /f/ and /θ/ fricative portions were spliced off from the recorded syllables. The fricative onset was defined by visually determining the location where the number of zero-crossing rapidly increases. The fricative offset was also visually determined by the intensity minimum before the onset of periodicity (Peterson & Lehiste, 1960). The spectrum of the six tokens is shown in Figure 2.

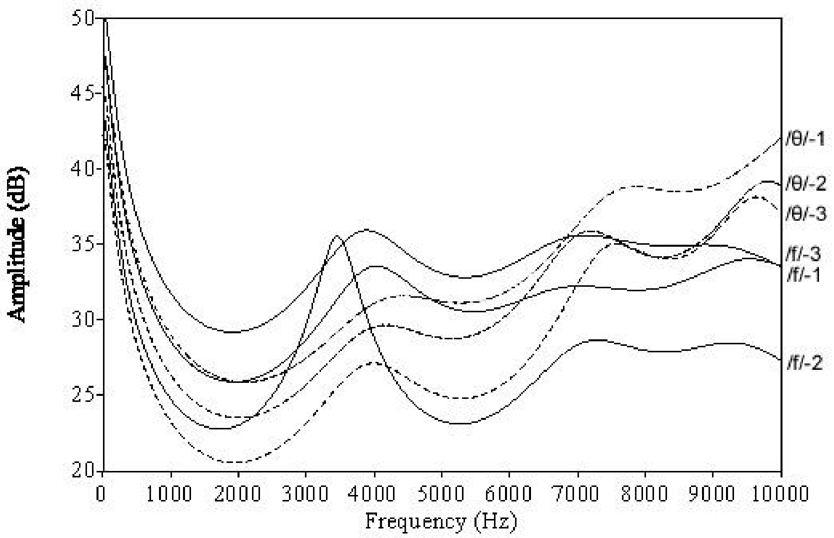

Figure 2. The spectrum of the six fricative tokens

From each fricative portion, four stimuli of different durations were created (Duration 100, Duration 75, Duration 50, and Duration 25). The durations were 100, 75, 50, and 25% of the natural fricative portions. Table 1 summarises 24 stimuli created by this procedure. For English native speakers, it is known that 40 ms is the shortest duration required for the correct identification of fricative portions, though the identification of /f/ and /θ/ is likely to require longer duration (Jongman, 1989). In the current experiment, the shortest duration was set around 40 ms which corresponds to the Duration

25. The purpose of creating the stimuli of different durations is to maximise the possibility of finding between-group differences in the perceptual sensitivity to the fricative portions. The effect of duration could have important implications in itself, but it will not be thoroughly discussed in the current study, because the primary interest is on differences between Japanese and English native speakers.

	Duration (ms)			
Token	100%	75%	50%	25%
/f/-1	168	126	84	42
/f/-2	162	122	81	41
/f/-3	201	151	101	50
/θ/-1	196	147	98	49
/θ/-2	162	122	81	41
/θ/-3	145	109	73	36

Table 1. Duration of 24 stimuli for Experiment 1

3.2.3 Procedure

The experiment was conducted in a quiet experiment room in the Department of Linguistic and English Language, the University of Edinburgh. The participants took part in the experiment in an individual booth equipped with a computer. The experimental software E-Prime was used to run the experiment and the sounds were played over headphones. The sound volume was adjusted to a comfortable level. The participants performed two tasks; (a) ABX discrimination task, and (b) similarity judgment task. Before the experiment started, the participants were informed that they would hear speech sounds in the experiment. The instructions for the tasks were displayed on the computer screen. The whole experiment took approximately 30 minutes. The participants were able to have a short break between the two tasks.

3.2.3.1 ABX discrimination task.

The fricative portions were paired with another fricative portion of the same duration to create a ABX discrimination pair. For example, if /f/-1 is A, B can be /θ/-1, /θ/-2, or /θ/-3 of the same duration. /f/-1 was not combined with the tokens of the same fricative, that is, /f/-2 and /f/-3. X is the same sound as either A or B. Thirty-six discrimination pairs were created for each duration set. The total number of discrimination pairs was 144 pairs (36 pairs × 4 durations). The participants were asked to decide if the final sound (X) was the same sound as either A or B. The interstimulus interval was set to 1000 ms and the intertrial interval was set to 1000 ms. The stimuli were blocked by the duration in order to prevent the participants from failing to hear the stimuli because the stimuli with the shorter durations have smaller intensity. Half of the participants heard the four blocks from the shortest to the longest, and the other half heard them in the reverse order. For each duration set, the participants heard the block of 36 stimuli once.

3.2.3.2 Similarity judgment task.

Seventy-two AB-AC similarity judgment pairs were created by combining the six tokens of Duration 100. In these AB-AC pairs, one pair consists of two tokens from the same sound (same-sound pair) and the other pair consists of two tokens from different sounds (different-sound pair). For example, if A is /f/-1 and B is /f/-2, C can be /θ/-1, /θ/-2, or /θ/-3. The 72 similarity judgment pairs were presented in randomised order. The participants were asked to choose a same-sound pair, hence a similar-sounding pair, between the two pairs. To choose the same-sound pair, participants were required not only to discriminate sounds but also to identify phonetically relevant similarities and differences in these sounds.

3.2.4 Results

3.2.4.1 ABX discrimination task.

Discrimination scores were calculated for each individual. The scores represent a percentage of the correct discrimination. A repeated measures analysis of variance (ANOVA) was carried out with Duration (4 levels) as a within-subjects variable and Native Language (2 levels) as a between-subjects variable. A dependent variable was the discrimination scores. There was a main effect of Duration, $F(3, 102) = 7.90$, $p < .01$, $\eta_p^2 = .19$. Pair-wise comparisons showed that the fricatives of Duration 75 (*M*

= 85.80, *SE* = 1.52), Duration 50 (*M* = 87.27, *SE* = 1.41) and Duration 25 (*M* = 86.73, *SE* = 1.59) were discriminated significantly better than those of Duration 100 (*M* = 80.25, *SE* = 1.21), $p < .01$. A main effect of Native Language was not significant, $F(1, 34) = 2.85$, $p > .05$. English native speakers (*M* = 86.77, *SE* = 1.47) showed the slightly higher discrimination score than Japanese native speakers (*M* = 83.27, *SE* = 1.47), but the difference did not reach the significant level. An interaction effect between Duration and Native Language was non-significant, $F(3, 102) = 2.16$, $p > .05$. In other words, the effect of Duration was not different in Japanese and English native speakers.

Although ANOVA showed the non-significant effect of Native Language, a series of independent *t* test showed some differences between the two groups. In Duration 50, English native speakers (*M* = 90.90, *SE* = 1.66) showed the higher discrimination score than Japanese native speakers (*M* = 83.64, *SE* = 2.29), $t(34) = -2.56$, $p < .05$ (two-tailed), $r = -.39$. In Duration 100, there was a tendency that English native speakers (*M* = 82.41, *SE* = 1.91) did better than Japanese native speakers (*M* = 78.09, *SE* = 1.50), $t(34) = -1.78$, $p = .084$ (two-tailed), $r = -.28$.

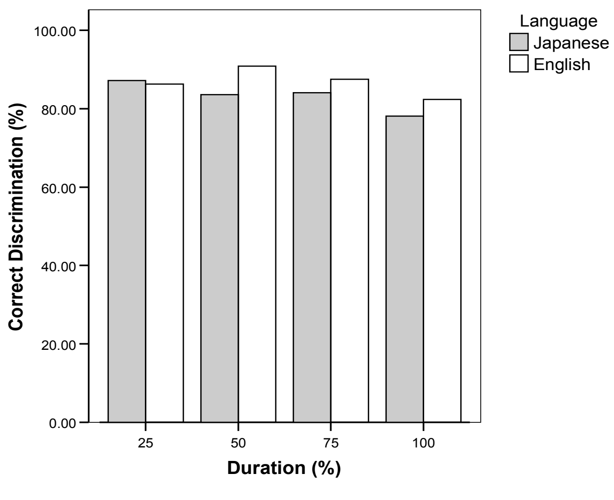

Figure 3. The correct discrimination of fricatives

3.2.4.2 Similarity judgment.

Judgment scores were calculated for each individual. The scores represent a percentage of the times that a same-sound pair was chosen. The judgment scores were submitted to an independent *t* test. English native speakers (M = 60.88, SE = 2.31) chose same-sound pairs more frequently than Japanese native speakers (M = 50.85, SE = 1.84). The difference between the two groups was highly significant, $t(34) = -3.39$, $p < .01$ (two-tailed), $r = -.49$. A binominal test showed that only two Japanese native speakers chose the same-sound pair above the chance, while half of the English native speakers chose the same-sound pair above the chance.

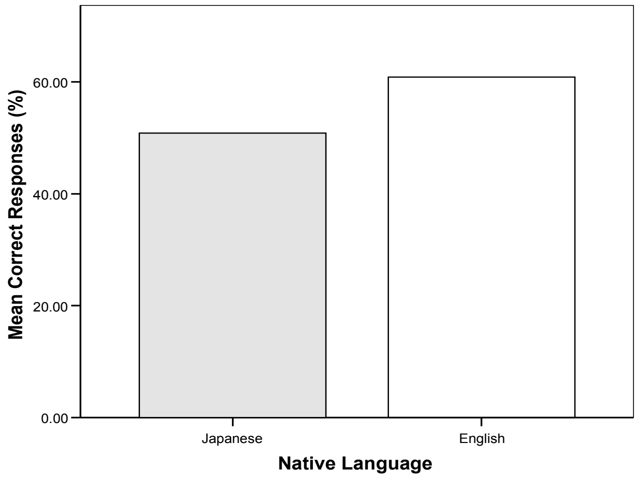

Figure 4. The correct similarity judgment of fricatives

3.2.5 Discussion

In the discrimination task, though the effect of language was not strong enough to be significant for the overall results, English native speakers discriminated /f/ and /θ/ better than Japanese native speakers in Duration 50 and possibly in Duration 100. Importantly, both Japanese and English native speakers did relatively well in the ABX discrimination task. Two factors seem to account for this result. First, there were various non-phonetic cues present in the stimuli created from natural speech. It was, therefore, possible that the participants discriminated /f/ and /θ/ based on non-phonetic cues, though they were informed that they would hear speech sounds. The fact that both groups did better in the shorter durations might indicate that the participants actually relied on non-phonetic cues to discriminate the stimuli. As

Jongman (1989) showed, phonetic identification becomes less accurate in the shorter durations. If the discrimination was largely based on the phonetically relevant cues in the current study, it is likely that the performance became worse in the shorter durations. Secondly, it is said that the discrimination of fricatives are generally high across and even within phonetic categories (Healy & Repp, 1982; Mann & Repp, 1980). In other words, discriminating /f/ and /θ/ fricatives might not be demanding whether they are native or non-native sounds.

However, regardless of these factors which might have facilitated the discrimination of /f/ and /θ/, the results showed that English native speakers did better than Japanese native speakers in Duration 50 and possibly in Duration 100. Thus, this result could support that non-native fricatives were actually more difficult to discriminate.

The effect of native language observed in the ABX discrimination task can be further supported by the results from the similarity judgment task. In the similarity judgment, the results showed that English native speakers were more likely to choose same-sound pairs. The result can be interpreted that English native speakers were more sensitive in detecting phonetically relevant similarities and dissimilarities, since they were likely to choose pairs consisting of the same sound regardless of various non-phonetic acoustic differences. On the other hand, Japanese native speakers were less likely to detect phonetically relevant cues which are necessary to categorise /f/ and /θ/.

Thus, the results from Experiment 1 showed that Japanese native speakers did not have as accurate a perception as English native speakers for the English /f/ and /θ/ fricatives. However, while English native speakers performed significantly better than Japanese native speakers, their similarity judgment score (60.88%) showed that the fricative portions were not enough for them to reliably categorise /f/ and /θ/. In other words, English native speakers also need formant transitions in the adjacent vowel in order to reliably perceive the difference between /f/ and /θ/.

3.3 Experiment 2: The Perception of Formant Transitions

The purpose of Experiment 2 was to examine if Japanese native speakers show the native-like perception of the formant transitions following /f/ and /θ/.

3.3.1 Participants

The same participants as Experiment 1 took part in Experiment 2 approximately 2 weeks after Experiment 1. Experiments 2 and 3 were carried out during the same experimental session. The order of the two experiments was counter-balanced across the participants.

3.3.2 Materials

3.3.2.1 A synthetic vowel continuum for the ABX identification task.

Seven synthetic vowels /ɑ/ with varying formant transitions were created by Sensyn Laboratory Speech Synthesiser (Sensimetrics Org.) based on a Klatt cascade/parallel formant synthesiser (Klatt, 1980). The vowel portions were synthesised at 20 kHz sampling rate which is the maximum value in Sensyn. Figure 5 shows stylised spectrograms of the seven stimuli with varying formant transitions.

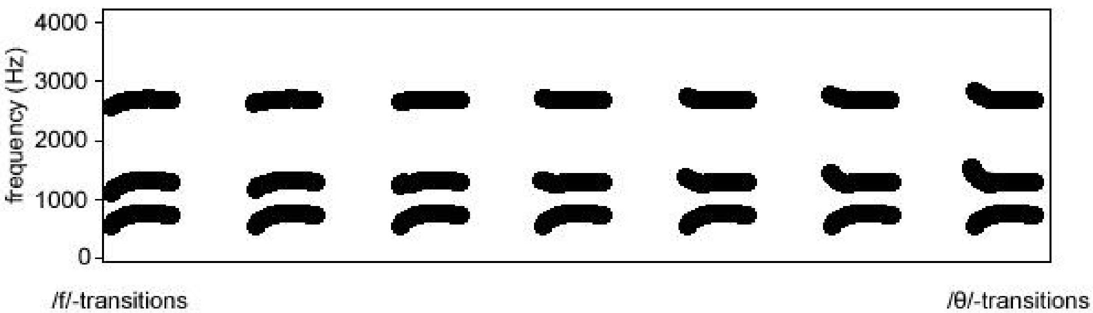

Figure 5. Stylised spectrograms of the synthetic vowel continuum

Various parameters were determined by examining the natural syllables /fɑ/ and /θɑ/ from which fricative portions for Experiment 1 were created. The duration was set to 500 ms. F0 falls from 250 Hz to 190 Hz within the first 300 ms. F1 transitions occurred within the first 100 ms from 580 Hz to 770 Hz. The two endpoints of the continuum had F2 and F3 transitions appropriate either for /f/ or /θ/. The F2 onset value was varied in 78 Hz steps from 1130 Hz (appropriate for /f/) to 1598 Hz (appropriate for /θ/). The steady-state value of F2 was 1300 Hz. The onset of F3 value was varied in 48 Hz steps from 2610 Hz (appropriate for /f/) to 2898 Hz (appropriate for /θ/). The steady-state value of F3 was 2750 Hz. The F2 and F3 transitions occurred within the first 100 ms (see Appendix for F2 and F3 onset values for the synthesis). After the synthesis, the sampling rate of synthetic vowels was converted to 22.05 kHz for the presentation on E-Prime and the amplitude was normalised based on the single peak amplitude using Adobe Audition.

3.3.2.2 Natural vowels for the similarity judgment task.

Six /ɑ/ tokens were spliced off from the natural speech used to create the stimuli in Experiment 1. Among the six tokens, three tokens had formant transitions appropriate for /f/ and the other three had those appropriated for /θ/. The pitch contour was manipulated using Praat so that F0 falls approximately from 250 Hz to 190 Hz within the first 60% of the total vowel duration. The purpose of the pitch manipulation was to prevent the participants from relying on pitch contour in the similarity judgment task. The amplitude was normalised after changing the pitch contour.

3.3.3 Procedure

The venue and the experimental equipments were the same as Experiment 1. The participants performed two tasks; (a) ABX identification task, and (b) similarity judgment task. Before the experiment started, the participants were informed that they would hear speech sounds in the experiment. The instructions for the tasks were displayed on the computer screen. The whole experiment took approximately 25 minutes. The participants were able to have a short break between the two tasks.

3.3.3.1 ABX identification task.

The participants heard a sequence of three sounds A, B, and X. A and B were the two endpoints of the continuum, that is, /ɑ/s which have the formant transitions appropriate for either /f/ or /θ/. X was one of the seven sounds from the continuum. The interstimulus interval was set to 1000 ms and the intertrial interval was set to 1000 ms. Participants were asked to decide if the final sound (X) was the same sound as either A or B. In this task, the participants were asked to identify sounds as /ɑ/ with either /f/-transitions or /θ/-transitions without explicitly labelling them. The participants first completed a practice session only with the endpoint stimuli as X. In the experimental session, they heard the block of 14 identification pairs (7 stimuli as X × 2 ordering of A and B) for five times. In total, the participants identified each stimulus for ten times.

3.3.3.2 Similarity judgment task.

This task followed the same procedure as the similarity judgment task conducted in Experiment 1. The stimuli were the six vowel portions instead of the six fricative portions.

3.3.4 Results

3.3.4.1 ABX identification task.

First, the perception of the two endpoint stimuli was examined. When X was either of the two endpoint stimuli, X was identical to either A or B. In other words, the participants simply had to perform a discrimination task for these stimuli. The percentage of the correct discrimination at the two endpoints were submitted to a repeated measures ANOVA with Endpoint (2 levels) as a within-subjects variable and Native Language (2 levels) as a between-subjects variable. There was no main effect of Endpoint $F(1, 34) = 0.009$, $p > .05$. The discrimination scores were similar at the /f/-transition end ($M = 72.22$, $SE = 3.79$) and the /θ/-transition end ($M = 72.50$, $SE = 4.05$). A main effect of Native Language was not significant, $F(1, 34) = 0.79$, $p > .05$. Japanese native speakers ($M = 71.39$, $SE = 5.12$) and English native speakers ($M = 73.33$, $SE = 5.12$) showed the same level of the discrimination score. There was no interaction between Endpoint and Native Language, $F(1, 34) = 0.24$, $p > .05$. The results showed that even the endpoint stimuli were correctly discriminated only around 70% level as a whole.

Secondly, in order to examine the overall identification function, the identification scores were calculated for each individual. The scores represent a percentage of /θ/-responses for each stimulus. The steepness and the location of category boundary were calculated by transforming the identification scores to probit scores (Finney, 1971). This procedure was commonly used by other studies (e.g., Mayo & Turk, 2004, 2005; Nittrouer & Miller, 1997). For this analysis, only the participants who correctly identified the endpoint stimuli above 80% were included. In the current study, this process excluded more than half of the participants in each group. Only eight Japanese and eight English participants were included in the analysis. There were some cases where the probit function did not show a good fit because of the one clear deviation near the endpoints of the S-shaped identification curve. In these cases, the original data point was modified in order to obtain the good fit of the curve. The results from these participants are shown in Figure 8.

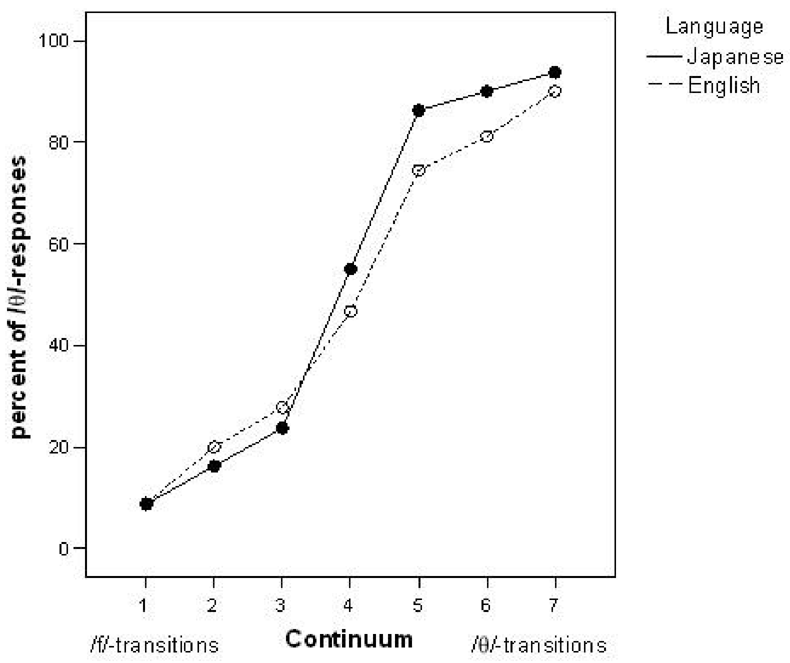

Figure 6. Identification functions for the vowel continuum

An independent *t* test showed that the steepness of the slope was not different between Japanese native speakers (M = 0.75, SE = 0.10) and English native speakers (M = 0.57, SE = 0.08), $t(14) = 1.41$, $p > .05$ (two-tailed). It was also showed that Japanese native speakers (M = 3.78, SE = 0.18) and English native speakers (M = 4.03 SE = 0.35) were not different in terms of the location of category boundary, $t(14) = -.63$, $p > .05$ (two-tailed).

3.3.4.2 Similarity judgment task.

The correct judgment scores were calculated for each individual. The scores represent a percentage of the times that a same-sound pair was chosen. One Japanese participant was a clear outlier and was excluded from the analysis. The result showed that Japanese native speakers (M = 88.15, SE = 2.69) chose the same-sound pair more often than English native speakers (M = 81.48, SE = 2.34), but an independent *t* test showed that the difference between the two groups was not statistically significant, $t(33) = 1.88$, $p > .05$ (two-tailed).

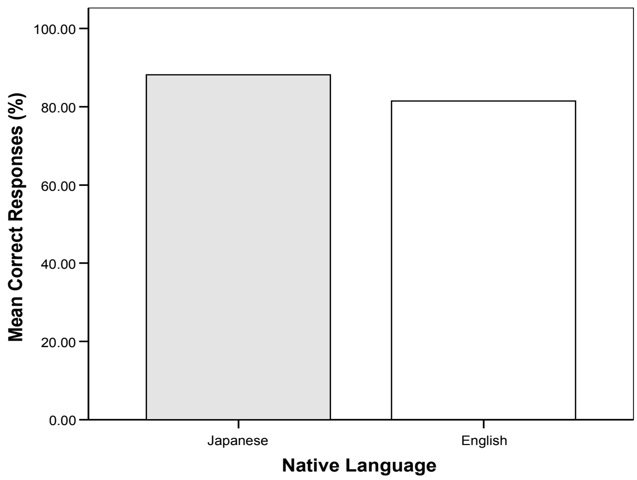

Figure 7. The correct similarity judgment of formant transitions

3.3.5 Discussion

In the ABX identification task, it was shown that Japanese and English native speakers had the same level of identifications for the formant transitions appropriate for /f/ and /θ/. There was no difference between the two groups in term of the steepness and the location of the category boundary between /f/ and /θ/. In the similarity judgment, the same tendency was held. Japanese native speakers chose same-sound pairs as often as English native speakers. Thus, the results of Experiment 2 showed that Japanese native speakers had the same degree of perceptual sensitivity to the formant transitions following /f/ and /θ/.

The results also showed that formant transitions alone were not sufficient to signal the /f/-/θ/ contrast. In the ABX identification task, the correct identification at the two endpoints reached only around 70% both in Japanese and English native speakers. This fact resulted in excluding more than half of the participants from the following analysis of the steepness of and the location of category boundary. It actually casts doubt on the experimental design used in the current study which assumed the perception above 80% at the two ends of the continuum. In the similarity judgment, the formant transitions were categorised better than in the ABX identification and the scores reached somewhere between 80 to 90%.

Experiments 1 and 2 examined the perception of individual cues contributing to the perception of the English /f/-/θ/ contrast. The results showed that Japanese native speakers were not able to categorise /f/ and /θ/ fricatives as accurately as English native speakers, while Japanese native speakers showed the same degree of categorisation for the formant transitions following /f/ and /θ/. In other words, while the informativeness of formant transitions stays the same for Japanese and English native speakers, the fricative cue is less informative for Japanese native speakers than for English native speakers. Thus, the relative informativeness of the two cues differs in the two groups.

According to the view that the informativeness of individual cues determines the weighting of these cues, the results from Experiments 1 and 2 predict different cue weightings between Japanese and English native speakers when they hear /fɑ/ and /θɑ/ syllables. It is expected that, in Japanese native speakers, the reduced informativeness of fricative cue results in the smaller weight assigned to the fricative. Accordingly, it is considered that Japanese native speakers assign greater weight to formant transitions than English native speakers do in the perception of /fɑ/ and /θɑ/ syllables. This hypothesis was tested in Experiment 3.

3.4 Experiment 3: The Perception of Fricative + Formant Transitions

The purpose of Experiment 3 was to examine how the two cues, fricative and formant transitions are weighted for the perception of /fɑ/ and /θɑ/ syllables. It was also examined if the syllable condition resulted in the native-like perception of /fɑ/ and /θɑ/ by Japanese native speakers.

3.4.1 Participants

The same participants as Experiments 1 and 2 took part in Experiment 3. Experiments 2 and 3 were carried out during the same experimental session. The order of the two experiments was counter-balanced across the participants.

3.4.2 Materials

3.4.2.1 Natural noise + synthetic vowel continua for the ABX identification task.

A /f/-continuum and a /θ/-continuum were created by combining one of the natural /f/ and /θ/ tokens (/f/-1 and /θ/-2) used in Experiment 1 and the synthetic vowels used in Experiment 2. Among 14 stimuli, one stimulus had /f/ fricative and formant transitions appropriate for /f/, and one had /θ/ fricative and formant transitions appropriate for /θ/. The selection of /f/-1 and /θ/-2 tokens among the natural tokens used in Experiment 1 was based on the following reasons. First, the duration of these two tokens was around 160 ms which is a typical value for /f/ and /θ/ (Jongman et al., 2000). Secondly, the closer investigation of the similarity judgment result in Experiment 1 showed that Japanese native speakers were not able to categorise /f/-1 and /θ/-2 as accurately as English native speakers. Thus, using these token ensured that the informativeness of fricative cue was different for the two groups. The spectrum of /f/-1 and /θ/-2 tokens are shown in Figure 8.

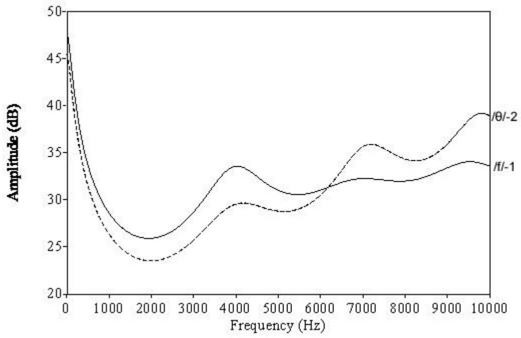

Figure 8.The spectrum of /f/-1 and /θ/-2

3.4.2.2 Natural syllables for the similarity judgment task.

These natural tokens for the similarity judgment were the syllables used to splice off the fricatives and vowels used in Experiments 1 and 2. The pitch contour was manipulated in the same way as Experiment 2.

3.4.3 Procedure

The venue and the experimental equipments were the same as Experiments 1 and 2. The participants performed two tasks; (a) ABX identification task, and (b) similarity judgment task. Before the experiment started, the participants were informed that they would hear speech sounds in the experiment. The instructions for the tasks were displayed on the computer screen. The whole experiment took approximately 25 minutes. The participants were able to have a short break between the two tasks.

3.4.3.1 ABX identification task.

The participants heard a sequence of three sounds A, B, and X. A and B were either /f/ + /f/-transitions or /θ/ + /θ/-transitions. In other words, A and B were /fɑ/ and /θɑ/ with non-conflicting cues. X was one of the sounds from 14 stimuli. The interstimulus interval was 1000 ms and the intertrial interval was 1000 ms. The participants were asked to decide if the final sound (X) was the same sound as either A or B. In this task, the listeners were asked to identify sounds as either /fɑ/ or /θɑ/ without explicitly labelling them. Participants first completed a practice session only with the /fɑ/ and /θɑ/ with non-conflicting cues as X to understand the task. In the experimental session, they heard the block of 28 stimuli (14 stimuli as X × 2 ordering of A and B) for five times. In total, the participants identified each stimulus for ten times.

3.4.3.2 Similarity judgment task.

This task followed the same procedure as the similarity judgment task conducted in Experiments 1 and 2. The stimuli consisted of the three tokens of /fɑ/ and the three tokens of /θɑ/.

3.4.4 Results

3.4.4.1 ABX identification task.

First, the perception of the two endpoint stimuli was examined. When X was either of the two endpoint stimuli, X was identical to either A or B. In other words, the participants simply had to perform a discrimination task for these stimuli. The percentage of the correct discrimination at the two endpoints was submitted to a repeated measures ANOVA with Endpoint (2 levels) as a within-subjects variable and Native Language (2 levels) as a between-subjects variable. There was no main effect of Endpoint, $F(1, 34) = 3.15$, $p > .05$. The discrimination scores were similar at the /f/ + /f/-transition end ($M = 89.72$, $SE = 2.03$) and the /θ/ + /θ/-transition end ($M = 85.83$, $SE = 2.57$). A main effect of Native Language was not significant, $F(1, 34) = .02$, $p > .05$. Japanese native speakers ($M = 88.06$, $SE = 2.88$) and English native speakers ($M = 87.50$, $SE = 2.88$) showed the same level of the discrimination score. There was no interaction between Endpoint and Native Language, $F(1, 34) = 1.03$, $p > .05$.

Secondly, in order to examine the overall identification function, the identification scores were calculated for each individual. The score represents a percentage of /θɑ/-responses for each stimulus. In order to examine the steepness and the categorical boundary of the identification function, the identification scores were transformed to probit scores (Finney, 1971) as in Experiment 2. For this analysis, only the participants who correctly identified the endpoint stimuli above 80% were included in the analysis. 13 Japanese and 12 English participants met this criterion. The results from these participants are shown in Figure 9.

The steepness of the slope was submitted to ANOVA with Fricative as a within-subjects variable and Native Language as a between-subjects variable. A dependent variable was the steepness of the slope calculated for each individual. The result showed that there was no effect of Fricative, $F(1, 23) = 3.36$, $p > .05$. The steepness of the slope was not different in the /f/-continuum ($M = 0.53$, $SE = 0.086$) and the /θ/-continuum ($M = 0.44$, $SE = 0.081$). A main effect of Language was not significant, $F(1, 23) = 0.42$, $p > .05$. Japanese native speakers ($M = 0.53$, $SE = 0.096$) and English native speakers ($M = 0.44$, $SE = 0.10$) showed the similar steepness. An interaction of Fricative and Native Language was not significant, $F(1, 23) = 0.20$, $p > .05$.

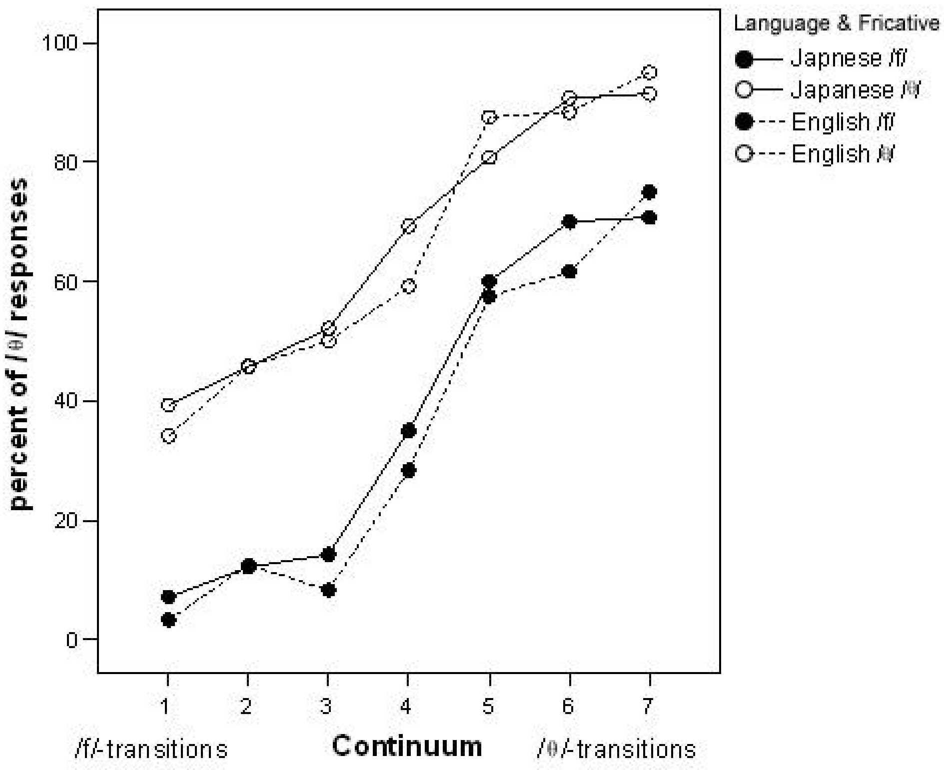

Figure 9. Identification functions for the /f/-continuum and the /θ/-continuum

The location of the category boundary was also submitted to ANOVA with

Fricative as a within-subjects variable and Native Language as a between-subjects variable. A dependent variable was the location of the category boundary calculated for each individual. The result showed that there was a main effect of Fricative, $F(1, 23) = 21.12$, $p < .01$, $\eta_p^2 = .48$. The location of category boundary was closer to the /f/-transition end for the /f/-continuum ($M = 5.34$, $SE = 0.50$) and was closer to the /θ/-transition end for the /θ/-continuum ($M = 1.96$, $SE = 0.45$). A main effect of Native Language was not significant, $F(1, 23) = 0.43$, $p > .05$. Japanese native speakers ($M = 3.90$, $SE = 0.42$) and English native speakers ($M = 3.41$, $SE = 0.44$) had the category boundary in the similar location. An interaction of Fricative and Native Language was not significant, $F(1, 23) = 0.45$, $p > .05$.

3.4.4.2 Similarity judgment task.

Correct judgment scores were calculated for each individual. The scores represent a percentage of the times that a same-sound pair was chosen. The judgment scores were submitted to an independent *t* test. The result showed that Japanese native speakers ($M = 96.22$, $SE = 0.91$) chose the same-sound pair more often than English native speakers ($M = 84.57$, $SE = 2.04$), $t(34) = 5.22$, $p <. 01$ (two-tailed), $r = .62$. The effect size indicated that the effect of native language was large.

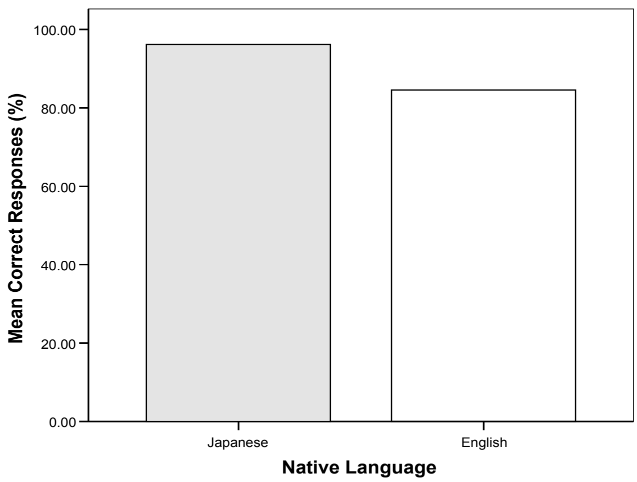

Figure 10. The correct similarity judgment of CV syllables

3.4.5 Discussion

According to the result from Experiments 1 and 2, it was expected that Japanese native speakers assigned greater weight to the formant transitions in the perception of the English /fɑ/-/θɑ/ contrast. However, the result of the ABX identification in Experiment 3 did not show any cue weighting differences between Japanese and English native speakers. The implications of this result will be discussed in the following chapter of general discussion.

The result of the similarity judgment task was somewhat puzzling. Japanese native speakers were likely to categorise /fɑ/ and /θɑ/ more sharply than English native speakers. This result showed that Japanese native speakers were able to reliably categorise the English /f/ and /θ/ in the syllable context, but it was not expected that the performance of Japanese native speakers exceeded that of English native speakers. The possible reason for this result will also be discussed in the following chapter of general discussion.

3.5 Summary

The results of the three experiments are summarised in Table 2.

	Tasks	
Cues	ABX discrimination/identification	Similarity Judgment
Fricative	English did slightly better	English did better
Formant transitions	No difference	No difference
Syllable	No difference	Japanese did better

Table 2. The results of three experiments

In the perception of fricatives, Japanese native speakers seem to be less able to categorise /f/ and /θ/ than English native speakers. In the perception of formant transitions following /f/ and /θ/, no difference was observed between Japanese and English native speakers. Finally, in the perception of /fɑ/-/θɑ/ syllables, while the ABX identification task showed no between-group differences, the similarity judgment indicated that Japanese native speakers categorised /fɑ/-/θɑ/ syllables more accurately than English native speakers.

4 General Discussion

Chapter 4 discusses the experimental results in light of the initial research questions. The discussion is also extended to various implications for future research.

4.1 The Perception of Individual Cues by L2 Learners

The Japanese participants' performance in the CV syllable condition showed that they could perceive the English /fɑ/-/θɑ/ contrast as accurately as, or even more accurately than, English native speakers. Then, the question is, as stated in the research question 1, the perception of the individual cues.

1. Does a good perception of L2 consonantal place contrasts in CV syllables indicate the native-like perception of individual cues?

The results from the three experiments seem to suggest that the native-like syllable perception does not indicate the native-like perception of individual cues. While Japanese native speakers perceived the English /fɑ/-/θɑ/ contrast as accurately as, or even more accurately than, English native speakers, Japanese native speaker did worse than English native speakers in the perception of /f/ and /θ/ fricatives. The English /f/-/θ/ contrast is known to be confusing even for English native speakers (Jongman, Wang, & Sereno, 2000), but it is also claimed that English native speakers can identify /f/ and /θ/ isolated fricatives to some extent (Jongman, 1989). In the current study, English native speakers were more likely to show the above-chance performance in the similarity judgment of fricatives. In other words, they were able to detect phonetically relevant similarities and differences between these two fricatives to some extent. As the ABX discrimination task showed, Japanese native speakers also heard the differences between /f/ and /θ/, but their chance-level performance in the similarity judgment indicates that Japanese native speakers were not able to decide which differences were phonetically relevant. This finding is particularly interesting in that the Japanese participants in the current study were advanced learners of English. Thus, the current study seems to suggest that Japanese native speakers were not able to perceive the English /f/-/θ/ fricative contrasts as accurately as English native speakers.

On the other hand, there was no difference between the two groups in the perception of /f/-/θ/ formant transitions. The result of Experiment 2 showed that Japanese native speakers performed as accurately as English native speakers both in the ABX identification and the similarity judgment tasks. Acoustically, /f/-transitions and /θ/-transitions are significantly different in terms of the direction of F2 and F3 movement; both F2 and F3 show rising transitions after /f/, while both F2 and F3 show falling transitions after /θ/. These different directions of formant transitions separate the labial and dental range of articulation not only in fricatives but also in stops. In fact, it is claimed that vowels with /f/-transitions or /θ/-transitions are often heard as stop-vowel syllables, if the listeners are not informed that the preceding consonant is a fricative (Carden et al., 1981). It is, therefore, likely that both Japanese and English native speakers in the current study heard the vowels with /f/-transitions and /θ/-transitions as /bɑ/ and /dɑ/ syllables. As Japanese also has /b/-/d/ contrast, it could explain why Japanese native speakers did as accurately as English native speakers in the current study.

However, this issue actually casts doubt on the experimental design used in the current study, especially in the ABX identification task, to assess the perceptual sensitivity to formant transitions as a cue for fricative place contrasts. Carden and colleagues (1981) created a vowel continuum with varying formant transitions which is similar to that used in the ABX identification task in Experiment 2 of the current study. In their study, half of the English-speaking participants performed a /b/-/d/ identification task, and the other half performed a /f/-/θ/ identification task to the same continuum. The result showed that the identification functions differed according to the identification-pair given to each group.

In the current study, the participants were simply instructed to identify the tokens from the continuum as either of the endpoint stimuli in order to avoid the labelling of the stimuli. Considering that vowels with /f/-transitions and /θ/-transitions are often heard as stop-vowel syllables (Carden et al., 1981), it is likely that both Japanese and English native speakers actually performed a /b/-/d/ identification task. Thus, it might not have been an appropriate method to assess the perception of formant transitions as a cue for the fricative place contrast. As Carden and colleagues (1981) showed, the performance must have been different, at least in English native speakers, if they were informed that they should identify the place of articulation of fricatives. Although it is not clear if this possible methodological problem concealed differences between Japanese and English native speakers, future studies should consider more appropriate experimental design to assess the perception of formant transition cues.

In summary, though there are some methodological issues to be considered especially with the perception of formant transitions, the current study at least seems to show that a good perception of L2 consonantal place contrasts in CV syllables does not indicate the native-like perception of individual cues. This finding is interesting in that difference between L2 learners and native speakers might not be found if the perceptual accuracy is tested only in syllable conditions. Further investigations of this issue could contribute to practical aspects of L2 speech perception study, such as perceptual training for L2 sound contrasts.

4.2 Informativeness and Weighting of Acoustic Cues

As Japanese native speakers showed different sensitivity to one of the cues involved in the perception of the English /fɑ/-/θɑ/ contrast, the next issue is how it affects the weighting of the two cues in the syllable perception.

2. If L2 learners show the different perceptual sensitivity for individual cues from native speakers, does it affect the weighting of these cues?

It is considered that native speakers weigh individual cues according to their informativeness for sound categorisation so that the reliable category perception can be achieved (e.g., Fischer-Jørgensen, 1972; Nittrouer & Miller, 1997). In the current study, as Japanese native speakers showed lower sensitivity to fricatives, it was likely that the relative informativeness of the fricatives became lower for Japanese native speakers. As a result, it was expected that Japanese native speakers assigned greater weight to formant transitions than English native speakers did.

However, the experimental result showed that Japanese and English native speakers did not have any cue weighting differences, even if the informativeness of fricatives was likely to be reduced in Japanese native speakers. There are three possible reasons for why there was no cue weighting difference between Japanese and English native speakers. The first two possibilities hold the assumption that cue weighting differences should be predicted by the informativeness of each cue. In other words, they attribute the result of the current study to various methodological and experimental factors. On the other hand, the third account casts doubt on the role of context-dependent informativeness in determining the cue weighting. The following two sections discuss these possibilities.

4.2.1 The Degree of Informativeness

It is possible that fricatives were informative enough for Japanese native speakers to assign the same perceptual weight as English native speakers did. The following two reasons seem to be possible in the current context.

First, though Japanese native speakers showed lower perceptual sensitivity to fricatives, it might be that their perception of fricatives was not low enough to reduce the relative informativeness of the fricatives and to cause cue weighting differences. That is, Japanese native speakers did relatively well in the discrimination task for /f/-/θ/ fricatives. This level of perceptual sensitivity might be enough to assign a certain weight to the fricatives in the perception of the /fɑ/-/θɑ/ contrast. Nittrouer and Lowenstein (2007) argue that it is extremely difficult to figure out how sensitive listeners should be in order to weigh a particular acoustic cue for phonetic decisions. In the current study, English native speakers were more sensitive than Japanese native speakers in terms of the perception of isolated fricatives and the difference was statistically significant. However, it does not show whether this difference is large enough to affect the way of cue weighting. Thus, Japanese native speakers might have been sensitive enough to weigh fricatives as heavily as English native speakers.

Secondly, a cause might has resided in the stimulus creation. Two factors seem to have increased the informativeness of the fricatives in the CV syllables. The first factor is dental realisation of /θ/ in the current study. For example, Nittrouer (2002) used the natural /f/ and /θ/ fricatives whose spectrum is highly similar and maximised the role of formant transitions in her cue weighting study of /fa/-/θa/ and /fu/-/θu/ in American English. However, English-speaking adults still weighted the fricatives to some extent. In the current study, /f/ and /θ/ fricatives used to create the stimuli were acoustically much more distinct than those used by Nittrouer (2002). /θ/ in Southern British Standard English is realised as dental rather than interdental (Ladefoged & Maddieson, 1996) and /f/ is more distinct from dental /θ/ than from interdental /θ/ (Jones, 2005). As a result, it is likely that the fricatives in the current study were more informative than those in the study by Nittrouer (2002).

The second factor is the combination of natural fricatives and synthetic vowels. In the cue weighting study of /sa/-/ʃa/ and /su/-/ʃu/ by Nittrouer (2002), synthetic /s/ and /ʃ/ fricatives were used to maximise the role of formant transitions since the natural fricatives of /s/ and /ʃ/ elicited the response exclusively relying on the fricatives. This fact implies that natural stimuli are much more informative than synthetic stimuli in general. In the current study, the synthetic vowels followed the natural fricatives. This synthetic-natural asymmetry might have increased the relative informativeness of the fricatives in contrast to the synthetic vowels. Thus, it is possible that, in the current study, the role of fricatives was enhanced by (a) the relatively distinct spectra of the /f/ and /θ/ fricatives, and (b) the natural fricatives in contrast to the synthetic vowels.

These two possible explanations postulate that cue weighting can be observed when the informativeness of fricatives is further reduced. In order to examine if the above stated factors actually affected the result of the current study, the experiment can be designed so that the fricatives become less informative, for example by creating stimuli by using /f/ and /θ/ fricatives whose spectrum is highly similar or by using synthetic /f/ and /θ/ fricatives. If it is shown that the cue weighting exists between the two groups, it could be concluded that the different relative informativeness in the two groups affects the weighting of the cues involved in fricative place contrasts.

4.2.2 Local Informativeness and Global Salience

Another possible reason why there was no cue weighting difference between Japanese and English native speakers is that cue weighting might not always be predicted from local informativeness of the relevant cues. Holt and Lotto (2006), for example, examined cue weighting mechanism using a perceptual space defined by two acoustic dimensions, carrier frequency (CF) and modular frequency (MF), and they found that CF was assigned greater weight even when the informativeness of CF and MF was equated. One of the possible interpretations for this result was that CF is globally more informative, in other words more salient, than MF, if the informativeness of CF and MF is considered within the whole perceptual experience of the perceivers (Holt & Lotto, 2006). Thus, local informativeness does not already predict the cue weighting in non-speech perception.

Although the study with non-speech stimuli should be cautiously interpreted when it is applied to speech perception studies, the distinction between local informativeness and global salience may offer some implications for L2 speech perception study. That is, while the local informativeness is determined in a specific L2 context, the global salience might be determined by both L1 and L2 experience. For example, in the current study, the local informativeness of the fricatives was likely to be reduced in Japanese native speakers. However, their experience with Japanese might have increased the salience of the fricatives. The Japanese fricative inventory does not have confusing place contrasts such as /f/-/θ/ or three-way sibilant contrasts. It is, then, likely that in many cases fricatives are much more informative than formant transitions in the perception of Japanese fricative place contrasts. If this L1 experience affects the global salience of fricatives, the reduced local informativeness of English /f/ and /θ/ might somehow be compensated by their global salience. As a result, the weight assigned to the fricative in the English /fɑ/-/θɑ/ perception cannot be predicted from the local informativeness of each cue, because the cue weighting is determined by the complex interaction between the local informativeness and the global salience.

This argument is highly speculative and the current study does not offer any direct evidence. However, it would probably be important for the future research to consider the global salience of acoustic cues determined by both L1 and L2 experience.

4.3 Individual Cue Perception and Configural Perception

The final section discusses the alternative approach to the current study. The current study attempted to add new insights into the study of L2 speech perception by examining the perception of individual acoustic cues and weighting of these cues. Although this approach has revealed various interesting phenomena in L2 speech perception, alternative approach is also possible.

The perception of the English /r/-/l/ is a good example to illustrate this issue. Iverson and colleagues (2003) claim that Japanese native speakers are not able to reliably perceive the English /r/-/l/ contrast because they are not sensitive to the changes in F3. However, Vallabha (2006) found that Japanese native speakers were not necessarily insensitive to the changes in F3 itself, and their insensitivity to F3 was due to a specific configuration of relevant cues in the English /r/-/l/ contrast, such as the duration of the syllables and the transitions of other formants. In his study, the same F3 changes were embedded in (a) the English /ra/-/la/ acoustic configuration, and (b) the Japanese /ra/ (apico-alveolar tap) acoustic configuration, and the perception of varying F3 was tested. The result showed that English native speakers detected the changes in F3 much better than Japanese native speaker in the English /ra/-/la/ configuration. However, when the same F3 change was presented in the Japanese /ra/ configuration, the performance of English speakers worsened while that of Japanese native speakers was improved. Based on this result, Vallabha (2006) claimed that L2 speech perception was better studied by a configural approach, rather than looking into the perception of isolated cues.

Although the current study was not designed to compare these two approaches, the experimental result might imply that the configural approach is very useful. In Experiment 3, Japanese native speakers performed better than English native speakers in the similarity judgment of the /fɑ/-/θɑ/ contrast. While English native speakers chose the same-sound pairs 85% of the time, Japanese native speakers chose them 96% of the time. It should, then, be asked why Japanese native speakers did much better than English native speakers in the syllable condition. It is rather unlikely that the quality of the stimuli caused the poorer performance of English native speakers, since the identity of the stimuli was checked by native speakers of Southern British Standard English prior to the experiments.

It seems difficult to predict Japanese participants' better performance based on their performance in the individual cue conditions. Japanese native speakers did worse than English native speakers in the perception of isolated fricatives and the performance of the two groups did not show any difference in the perception of formant transitions. Thus, there seems to be no source which enabled Japanese native speakers to perform better than English native speakers in the perception of the CV syllables.

On the other hand, if the finding by Vallabha (2006) is applied to the current study, it might be that the fricatives or the formant transitions became a powerful place cue for Japanese native speakers only in the CV configuration. Thus, this speculation raises the question of whether the perception of acoustic cues should be examined in isolation or in specific acoustic configurations. Although the current study was based on the assumption that the perception of isolated individual cues matters, future studies should give full consideration to the alternative possibility.

5 Conclusion

The current study showed that the native-like perception of L2 place contrasts in CV syllables did not indicate the native-like perception of individual acoustic cues. This finding is in line with other L2 speech perception studies which found that the difference between native speakers and L2 learners could be found only when small details are examined. However, the current study did not manage to link this finding to cue weighting differences. Various possibilities were discussed. The result might be attributed to the experimental and methodological issues of the current study and the difference between Japanese and English native speakers might have been concealed. However, it is also possible that the result actually reflected important issues related to cue weighting mechanism in general. Future studies are necessary to pin down these remaining problems.

It is also important to give full consideration to the principle assumption of the current study. There were some indications that L2 perception might have been studied better by examining the perception of acoustic cues in a specific acoustic configuration rather than in an isolated condition. The investigations of this issue could contribute not only to L2 speech perception study but also to speech perception in general.

A wide scope of remaining issues should, therefore, be considered in future research. Methodological issues might need to be reviewed to discover finer differences between L2 learners and native speakers, and theoretical issues also need to be attended to in detail. In L2 speech perception study, it is also important to replicate the findings in different combinations of native and L2 languages, in order to generalise the findings.

References

Aoyama, K. (2003). Perception of syllable-initial and syllable-final nasals in English by Korean and Japanese speakers. *Second Language Research*, *19*(3), 251-265.

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of Optical Society of America*, *20*(7), 1391-1397.

Behrens, S. J., & Blumstein, S. E. (1988). Acoustic characteristics of English voice-less fricatives: A descriptive analysis. *Journal of Phonetics*, *16*, 295-298.

Best, C. T. (1990). Adult perception of nonnative contrasts differing in assimilation to native phonological categories. *Journal of the Acoustical Society of America*, *88*, S177.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp.171-204). Baltimore: York Press.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross language perception of approximants. *Journal of Phonetics*, *20*, 305-330.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In B. Ocke-Schwen, & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp.13-34). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Bladon, A., Clark, C., & Mickey, K. (1987). Production and perception of sibilant fricatives: Shona data. *Journal of the International Phonetic Association*, *17*(1), 39-65.

Blumstein, S. E. & Stevens, K. N. (1979). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America*, *67*, 648-662.

Bonneau, A., Djezzar, L., & Laprie, Y. (1996). Perception of the place of articulation of French stop bursts. *Journal of the Acoustical Society of America*, *100*(1), 555-564.

Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *Journal of the Acoustical Society of America*, *117*(6), 3890-3910.

Byrd, D. (1993). Preliminary results on speaker-dependent variation in the TIMIT database. *Journal of the Acoustical Society of America*, *92*, 593-596.

Carden, G., Levitt, A., Jusczyk, P. W., & Walley, A. (1981). Evidence for phonetic processing of cues to place of articulation: Perceived manner affects perceived place. *Perception & Psychophysics, 29*(1), 26-36.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech Perception. *Annual Review of Psychology*, *55*, 149-179.

Dufour, S., Nguyen, N., & Frauenfelder, U. H. (2007). The perception of phonemic contrasts in a non-native dialect. *Journal of the Acoustical Society of America*, *121*(4), 131-136.

Finney, D. J. (1971). *Probit analysis*. Cambridge: Cambridge University Press.

Fischer-Jørgensen, E. (1972). Tape-cutting experiments with Danish stop consonants in initial position. *Annual Report, Institute of Phonetics, University of Copenhagen*, *6*, 104-168.

Flege, J. E. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp.233-277). Baltimore: York Press.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America*, *107*(5), 2711-2724.

Hallé, P. A., Best, C. T., & Levitt, A. (1999). Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics*, *27*, 281-306.

Harnsberger, J. D. (2001). On the relationship between identification and discrimination of non-native nasal consonants. *Journal of the Acoustical Society of*

America, *110*(1), 489-503.

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, *1*, 1-7.

Healy, A. F., & Repp, B. H. (1982). Context independence and phonetic mediation in categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(1), 68-80.

Heinz, J. M., & Stevens, K. N. (1961). On the Properties of Voiceless Fricative Consonants. *Journal of the Acoustical Society of America*, *33*(5), 589-596.

Hicks, C. B., & Ohde, R. N. (2005). Developmental role of static, dynamic, and contextual cues in speech perception. *Journal of Speech, Language, and Hearing Research*, *48*, 960-974.

Hirai, S., Yasu, K., Arai, T., & Iitaka, K. (2005). Perceptual weighting of syllable-initial fricatives for native Japanese adults and for children with persistent developmental articulation disorders. *Sophia linguistica*, *53*, 49-76.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical So ciety of America*, *119* (5), 3059-2071.

Hughes, G. W., & Halle, M. (1956). Spectral properties of fricative consonants. *Jour nal of the Acoustical Society of America*, *28*, 303-310.

Ingram, J. C. L., & Park, S. –G. (1998). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, *25*, 343-370.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Tekkermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57.

Jones, M. J. (2005). An experimental acoustic study of dental and interdental non-sibilant fricatives in the speech of a single speaker. *Cambridge Occasional Papers in Linguistics*, *2*, 109-121.

Jongman, A. (1989). Duration of frication noise required for identification of English fricatives. *Journal of the Acoustical Society of America*, *85*(4), 1718-1725.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, *108*(3), 1252-1263.

Jongman, A., Wang, Y., & Kim, B. H. (2003). Contributions of semantic and facial information to perception of nonsibilant fricatives. *Journal of Speech, Language, and Hearing Research*, *46*, 1-11.

Jongman, A., Wang, Y., & Sereno, J. A. (2000). Acoustic and perceptual properties of English fricatives. *Proceedings of the 6th International Conference on Spoken Language Processing* , *I*, 536-539.

Klatt, D. (1980). Software for a cascade /parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*, 971-995.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet ef fect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*, 93-107.

Kurowski, K., & Blumstein, S. E. (1984). Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants. *Journal of the Acoustical Society of America*, *76*(2), 383-390.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Lambacher, S., Martens, W., Nelson, B., & Berman, J. (2001). Identification of English voiceless fricatives by Japanese listeners: The influence of vowel context on sensitivity and response bias. *Acoustical Science and Technology*, *22*(5), 334-343.

LaRiviere, C., Winitz, H., & Herriman, E. (1975). The distribution of perceptual cues in English prevocalic fricatives. *Journal of Speech and Hearing Research*, *18*, 613-622.

Li, F., Edwards, J., & Beckman, M. E. (2007). Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese. *Proceedings of the XVIth In ternational Congress of Phonetic Sciences,* 6-10 August 2007, Saarbruecken.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signalling /b/ versus /p/ in trochees. *Language and Speech*, *29*, 384-422.

Lisker, L., & Abramson, A. S. (1970). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384-422.

Malécot, A. (1956). Acoustic cues for nasal consonants: An experimental study involving a tape-splicing technique. *Language*, *32*, 274-284.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, *28*, 213-228.

Mayo, C., & Turk, A. (2004). Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. *Journal of the Acoustical Society of America*, *115*(6), 3184-3194.

Mayo, C., & Turk, A. (2005). The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception. *Journal of the Acoustical Society of America*, *118*(3), 1730-1741.

Morrison, G. S. (2002). Japanese listeners' use of duration cues in the identification of English high front vowels. In J. Larson, & M. Paster (Eds.), *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society* (pp. 189-200). Berkeley, CA: Berkeley Linguistics Society.

Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America*, *112*(2), 711-719.

Nittrouer, S., & Lowenstein, J. H. (2007). Children's weighting strategies for word-final stop voicing are not explained by auditory sensitivities. *Journal of Speech, Language, and Hearing Research*, *50*, 58-73.

Nitteour, S., & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America*, *101*(4), 2253-2266.

Nowak, P. M. (2006). The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics*, *34*, 139-152.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, *32*, 693-703.

Repp, B. H. (1982). Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. *Psychological Bulletin*, *92*(1), 81-110.

Sensimetrics Org. (n.d.). *Sensyn: Spech Synthesizer Package*. Camridge, MA.

Smits, R., Ten Bosch, L., & René, C. (1996). Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. I. Perception experiment. *Journal of the Acoustical Society of America*, *100*(6). 3852-3862.

Stevens, P. (1960). Spectra of fricative noise in human speech. *Language and Speech*, *3*, 32-49.

Tingsababh, K., & Abramson, A. S. (1993). Thai. *Journal of the International Phonetic Association*, *23*, 24-28.

Tsukada, K. (2006). Cross-language perception of word-final stops in Thai and English. *Bilingualism: Language and Cognition*, *9*(3), 309-318.

Vallabha, G. K. (2006). The role of acoustic familiarity in the perception of L2 contrasts. *151st Meeting of the Acoustical Society of America*, June 5-9, Providence, RI.

Wagner, A., Emestus, M. & Cutler, A. (2006). Formant transitions in fricative identification: The role of native fricative inventory. *Journal of the Acoustical Society of America*, 120(4), 2267-2277.

Walley, A. C., & Carrell, T. D. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *73*(3), 1011-1022.

Wardrip-Fruin, C., & Peach, S. (1984). Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants. *Language and Speech*, *27*, 367-379.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behaviour and Development*, *7*, 49-63.

Appendix

F2 and F3 onset values for the synthetic vowel continuum

	Values	
Continuum	F2 onset	F3 onset
1 (/f/-transitions)	1130	2610
2	1208	2658
3	1286	2706
4	1364	2754
5	1442	2802
6	1520	2850
7 (/θ/-transitions)	1598	2898